Journal of Theoretical and Applied Mechanics, Sofia, Vol.54 (2024) pp. 242-252
GENERAL MECHANICS

ENHANCING NEURAL NETWORK EFFICIENCY IN AUTOMATED IMAGE ANALYSIS FOR THERMAL NONDESTRUCTIVE TESTING

DENYS STOROZHYK[1], ANATOLII PROTASOV[1*],
YURII KUTS[1], OLEKSANDR MURAVIOV[1], IULIIA LYSENKO[1],
YORDAN MIRCHEV[2]

[1]*National Technical University of Ukraine,*
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
[2]*Institute of Mechanics at the Bulgarian Academy of Sciences, Sofia, Bulgaria*

[Received: 31 August 2023. Accepted: 11 May 2024]

doi: https://doi.org/10.55787/jtams.24.54.2.242

ABSTRACT: The article is devoted to the study of the efficiency of using the image fusion method for data augmentation when training neural networks for thermal non-destructive testing. Using the example of training 75 neural networks with the AlexNet architecture, the dependence of image classification accuracy on the data augmentation method was analyzed. The effectiveness of generating artificial images using the method of data integration based on wavelet transformation was examined. The results of the analysis demonstrate the advantages of using image fusion as a method of data augmentation and enrichment to improve the efficiency of neural networks. The superiority of the data augmentation method based on image fusion using wavelet transform is proven.

KEY WORDS: image fusion, data augmentation, neural networks, data enrichment.

1 INTRODUCTION

Classifying objects by their thermal images is is a common practice in thermal non-destructive testing. In technical applications, this task is crucial for the integrity control of structural elements and the study of materials' thermophysical properties. It is also relevant for analyzing the operating modes of machines and mechanisms, where elements may heat up due to energy transformation. The related problems are solved with computer vision methods, with modern systems giving the best results thanks to the integration of artificial intelligence technologies. The availability of

*Corresponding author e-mail: slp_fce@uacg.bg

computational power and the efficiency of modern neural networks make it possible to use computer vision to solve a wide range of current problems [1,2].

One of the most important challenges in developing a neural network is the insufficient amount of data and the uneven distribution of data between classes in the training sample. A lack of data significantly reduces the efficiency of artificial intelligence systems and can lead to overfitting. In this case, the model may perform well with the training set but will not efficiently process new data. Irregularity between classes can occur in areas such as non-destructive testing (NDT) and diagnostics. This is due to the fact that in real production and operation conditions, product defects vary in nature, geometric dimensions, shape, and frequency of occurrence [3].

Increasing the sample size for training can be achieved by accessing necessary data from open sources or through additional empirical research. However, suitable training datasets are often not publicly available, and the budget and time for such studies are always limited [4]. Another way to expand the dataset for neural network training is through artificial methods, such as data augmentation [5, 6]. These methods are widely used in deep learning, especially in computer vision, to improve accuracy. Data augmentation involves artificially increasing the size of the training dataset to enhance neural network efficiency. Additionally, data augmentation can expand limited datasets, increase data diversity, and prevent overfitting in artificial intelligence systems.

2 Analysis of Recent Research and Publications

Data augmentation methods can be divided into traditional methods and advanced methods, such as 'black box' approaches using Generative Adversarial Networks (GANs). The traditional ones are divided into several groups: methods of affine image transformations (such as rotation, scaling, displacement), color modification (increasing contrast, changing white balance, sharpening the image), adding noise, using the random erasure method, and mixing images [4]. In addition to traditional methods, data augmentation based on Generative Adversarial Networks (GANs) is also widely used. GANs can generate new images by transferring textures (styles) from one image to another [7] and blending images [8]. While these new images closely resemble the originals, the application of each method can significantly impact the neural network's performance in tasks like classification or detection [9].

Optimal results are obtained by selecting combinations of augmentation methods tailored to specific tasks. Current research aims to enhance algorithms for the automated selection of these combined augmentation methods [10].

Paper [11] offers a comprehensive overview of image augmentation methods for deep learning. The authors categorize these methods into three groups: methods without models, model-based methods, and those based on optimization approaches.

Without model methods employ image processing techniques, while model-based methods utilize trained image generation models. Optimization-based methods aim to identify optimal operations or combinations thereof. In our view, model-based methods show the most promise.The "Sample Pairing" method proposed in [12] is based on combining two randomly selected images from the training set, which may belong to different classes. The image thus generated belongs to the class of the first selected image. The application of this method made it possible to reduce the first-level error rate by 4.5%. The "Mixup" method described in [13] shares a similar principle with "Sample Pairing", but differs in that the resulting images' class is determined by the original images' proportions. These methods are effective with small datasets, allowing the training sample size to expand to N^2, where N represents the number of elements in the initial sample. However, a drawback of these methods is the introduction of additional errors due to manipulation of class membership. For instance, authors often need to exclude augmented data from the sample during the final training epochs.The authors of [14] proposed two data augmentation methods: one based on discrete wavelet transform (DWT), and the other on a constant quality factor transformation (CQT). The data augmentation method based on DWT showed maximum performance in only one of the data sets. The algorithm was a random replacement of the transform wavelet coefficients by the coefficients from one of five randomly selected images of the same class. The proposed methods were tested on four data sets. Research results demonstrated the need to use additional approaches for data augmentation, including those based on deep learning, such as GANs, and on a much larger number of datasets.

The aim of this work is to study the effectiveness of applying the method of image mixing, namely, the image fusion for augmentation of the training sample in the problems of thermal non-destructive testing and diagnostics.

Image fusion involves combining multiple images of the same object or scene to enhance the information content of the resulting image [15]. This method differs from image mixing methods discussed earlier. Unlike mixing, image fusion does not blend images from different objects within the same class or alter the inherent characteristics of a single object in the fused image. Examples include integrating images from different modalities, varying focus levels, or taken at different times.

In recent studies [16], image changes resulting from the temperature fluctuations of heated objects over time and the natural hand tremors of operators have been explored. One of the primary advantages of image fusion methods is their ability to enrich data by enhancing information content [17].

3 Experimental Data Description

In current research, significant efforts are dedicated to improving NDT methods for the automatic detection and classification of defects within various materials' internal structures. Detecting such defects presents a complex challenge that directly impacts product and construction quality. Thermal testing stands out as a popular NDT method widely employed across diverse industries.

This experiment aims to evaluate the efficacy of fusing object thermograms to augment training data. We acquired 400 infrared (IR) images depicting internal defects categorized into four types: circle, square, triangle, and no defect. These samples featured artificial internal defects in getinaks materials, located 2 mm beneath the surface. Examples of the acquired thermograms are illustrated in Fig. 1.

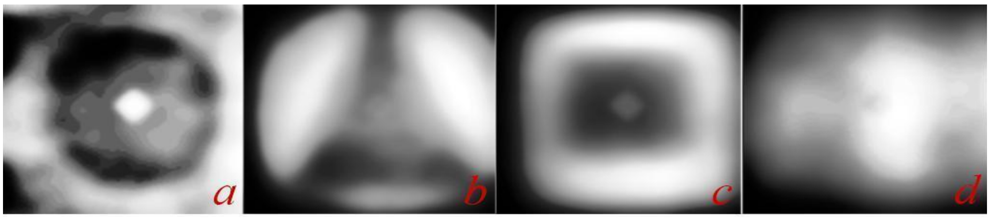

Fig. 1: Thermal images depicting defects from various categories: a, b, c – test samples with artificial defects in the form of a circle, triangle and square, respectively; d – test sample without defects

Each data category contains 100 images taken in series by the following procedure. The samples were heated with an IR electric heater, after that, images were taken by a thermal imager with a delay (several tens of seconds) and without a tripod. This procedure led to the presence of small differences between the images of the same series.

4 Implementation of the Experiment

The experiment involved a comparative analysis of several augmentation methods: noise addition, Sample Pairing, random replacement of DWT coefficients, and image fusion based on DWT [18]. Additionally, a control test was conducted without augmentation.

Augmentation methods are typically implemented in three common ways: expanding the dataset, adding data on-the-fly, and combining preliminary approaches.

Expanding the data set. The result of such augmentation is the creation of a new extended data set generated by random transformations. With this approach, a large training set is generated from a small one, but it is not possible to increase the model's ability to generalize. In addition, the implementation of such an approach requires significantly greater computing power.

Adding on-the-fly data. With this approach, the data is randomly replaced by generated data from one of the augmentation methods, but the total size of the training set is not increased. That is, the augmented data partially replaces the initial data.

For the implementation of this experiment, the on-the-fly data augmentation approach was chosen due to its advantages and common use.

One of the factors that influence the accuracy of the model during training is randomness in the formation of initial weights and the division of data into training and validation samples [19]. The first randomness is overcome by hard tuning the algorithm of a pseudo-random number generator. To overcome the second one, the principle of cross-validation (k-fold validation) is used, which is based on random splitting of the entire sample into k parts (Fig. 2). Consistently, each part plays the role of a validation sample, while the others act as a training one. Therefore, after one such cycle, k models will be obtained, which makes it possible to overcome the effect of "happy sampling", and allows application of statistical processing methods. Next cycles of cross-validation increase the neural network estimation accuracy. As part of the experiment, three repeated cycles were carried out with the division of data into five parts, which resulted in 15 models for each of the comparative categories.

In Fig. 2, the circles represent the data. The rectangles are the parts into which the sample is divided. At each iteration, the validation sample is changed and indicated by the blue rectangle, the training sample is indicated by the white rectangles.

Fig. 2: Principle of neural networks cross-validation.

The AlexNet structure was chosen as the architecture of the neural network. Training takes place over 30 epochs. AlexNet is a classic neural network architecture for computer vision systems that provides sufficient accuracy for solving simple problems (for example, determining the objects shape). At the same time, this architecture has a relatively small number of parameters (60 million) to be efficient with small training samples and does not require large computing power. The learning objective is to classify four types of thermal images according to four defect categories in infrared images. The accuracy of images classification (1) by neural networks was chosen as a quality metric. During the study, after each epoch, the accuracy of the variational sample was checked to keep the weights with the best accuracy

$$\text{accuracy} = \frac{N_{\text{true}}}{N}\,, \tag{1}$$

where N is the total number of images, N_{true} is the total number of correctly classified images. The obtained classification accuracies of the test sample for each trained neural network are recorded in the database for further analysis.

5 Analysis of the Results

The results of the experiment and assessment of the classification accuracy of the validation sample by neural networks, depending on the augmentation method, are shown in Fig. 3.

The lower boundary of the closed rectangular area corresponds to the first quartile, the upper boundary to the third quartile, and the central line to the median.

As can be seen from the diagrams, the nature of the classification accuracy distribution of the test sample is non-Gaussian. Due to the non-normal distribution of samples and the potential presence of outliers, the Kruskal-Wallis criterion [20] was chosen as a method for comparing these samples. It is a non-parametric criterion for testing the hypothesis that the medians of several samples are equal. For this criterion, the null hypothesis H0 is the absence of significant differences between the sets of classification accuracies.

The criterion statistics were determined by the formula:

$$H = \frac{12}{N\,(N-1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} 3\,(N+1)\,, \tag{2}$$

where k is the number of groups used in the comparison, N is the total sample size, n_i is the sample size for the i-th group, R_i is the total number of the i-th group ranks.

The value of this statistic was $H = 36.17$ for four degrees of freedom ($df = 4$); statistical significance (Sig.) was obtained close to zero. The significance obtained

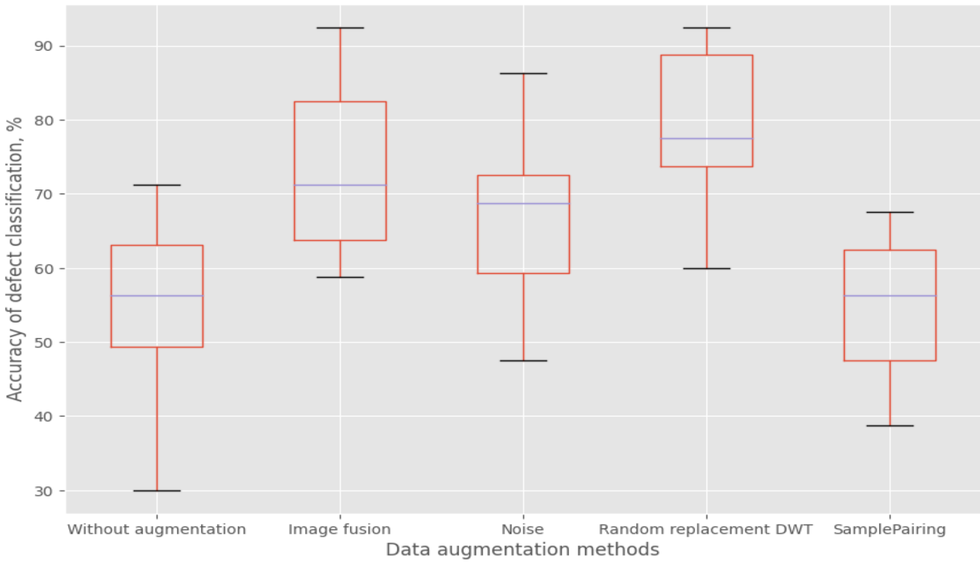

Fig. 3: Box diagrams of the validation sample classification accuracy depending on the type of the training sample augmentation.

level is less than the industry standard (p-value at the level of 0.05), so the null hypothesis H0 about the absence of significant differences between the samples is rejected. Therefore, the differences between the samples are significant. That is, the relationship between the augmentation method and classification accuracy is statistically proven.

Let's consider the results for each category in ascending order of median accuracy. The reference sample was trained without the augmentation data. Its median accuracy is 56%. The Sample Pairing method received a median classification accuracy value of 56%. We can conclude that this method almost did not improve the classification accuracy, although its minimum value is much higher than in the reference version. For the noisy method, the median was 68.7% accuracy. This augmentation method was chosen to provide reference accuracy characteristics when using standard augmentation methods.

When using image fusion, a median of 71.2% was obtained. The median accuracy increased by 2.5% compared to the noisy method and it increased by 15.2% compared to the case of classification without augmentation.

The median accuracy rate of 77.5% was obtained for the DWT random coefficient replacement method. In this case, an increase in the median accuracy by 5.3% was obtained compared to image fusion. However, if we look at Fig. 2, we will see

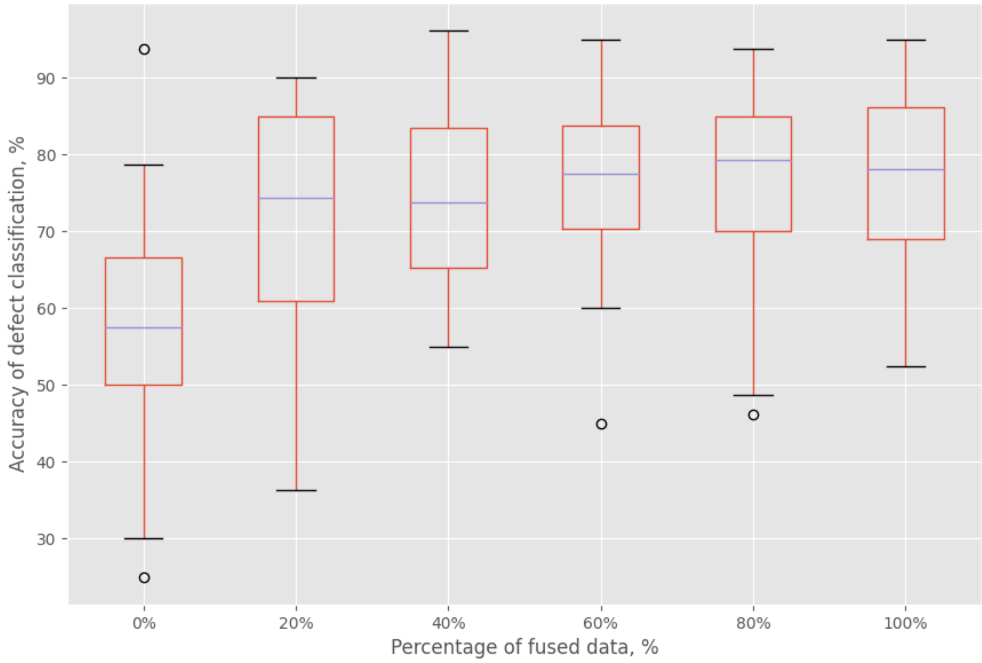

Fig. 4: Box plots of the classification accuracy of the validation set depending on the number of fusion images in the training set.

the accuracy rates of the two categories overlap almost completely, so it is worth checking whether there is a statistically significant difference between them. We use the Kruskal-Wallis criterion to do this, but now only for two samples. H-statistic is 2.24, statistical significance at the level of 0.14, which is significantly higher than the p-value of 0.05, so the null hypothesis about the absence of significant differences between the samples cannot be rejected. Therefore, there is no statistically significant difference between the accuracy samples of the two categories.

It is advisable to investigate the dependence of classification accuracy on the augmented images number in the training set for a more detailed analysis of the using image fusion possibilities as a data augmentation method. For this purpose, 30 neural networks of the AlexNet architecture were trained on training samples of the same size using the data cross-validation technique, but with a different percentage of fusion images.

The results of image classification accuracy depending on the relative amount of fusion data in the training sample are shown in Fig. 4.

The presence of 20% of fusion images from their total number in the training sample significantly increases the median value of the classification accuracy, which is 17% compared to a sample that does not contain fusion data at all. Subsequent increases in the relative number of fusion images in the training set lead to an increase in the classification accuracy within an additional 6% with little fluctuations. The best result was obtained by the training sample from 80% of fusion images, for which the increase in the median value of classification accuracy was 23%.

The received results demonstrate that in order to obtain a significant increase in classification accuracy due to the use of data augmentation by the method of image fusion, it is sufficient to have 20% of fusion data.

The data fusion method allows combining multimodal data and thus increases the information content of the resulting images [21]. In addition, this method has a simpler implementation algorithm compared to the DWT random coefficient replacement method, for example, data fusion method requires only two images to be processed, while the random coefficient replacement method requires six. Thus, the data fusion method can be used as a more effective tool for data augmentation and enrichment.

Conclusions

A statistical analysis of four augmentation methods in the problems of classifying thermal images was carried out. The analysis was conducted according to the k-fold validation method of neural networks and using the Kruskal-Wallis criterion. A significant advantage of methods based on wavelet transformation has been statistically proven. The use of these methods gives an increase in classification accuracy up to 21% compared to the base sample and up to 10% compared to the other best augmentation method studied.

The study examines how classification accuracy correlates with the proportion of fusion images in training samples. Results indicate that integrating 20% fusion images into the total training dataset enhances median image classification accuracy by up to 17%. Further increasing the proportion of fusion data can lead to accuracy improvements of up to 23%. This data augmentation method for neural networks demonstrates potential for automating the detection and classification of internal defects using thermal testing methods.

Future research should focus on assessing the effectiveness of data augmentation through image fusion across various modern neural network architectures and different training datasets, particularly in conjunction with alternative augmentation methods.

References

[1] V. Monfared (2017) Neural network based simulation of micro creeping fibrous composites SiC/Al6061 for plastic behavior. *Journal of Theoretical and Applied Mechanics, Sofia* **47** (1) 36-48.

[2] S. Khaoula, Y.M. Athmane, S. Belhadi, T. Mabrouki, S. Chihaoui (2023) Modeling of cutting force and power consumption using ANN and RSM methods in turning of AISI D3: comparative study and precision benefit. *Journal of Theoretical and Applied Mechanics, Sofia* **53** 49-65.

[3] R.M. Galagan, A.S. Momot (2017) The use of ART-2 neural network for processing information signals of non-destructive testing. Presented at *IEEE First Ukraine Conference on Electrical and Computer Engineering* (UKRCON).

[4] N.E. Khalifa, M. Loey, S. Mirjalili (2022) A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artificial Intelligence Review* **55**(3) 2351-2377.

[5] M. Giles (accessed Oct. 3, 2023) The GANfather: The man who's given machines the gift of imagination. https://www.technologyreview.com/.

[6] O. Chaikovskyi, A. Volokyta, A. Kyrianov, H. Loutskii (2021) Data augmentation method using generative adversarial networks. *Technical Sciences and Technology* **2**(24) 83-91.

[7] Y. Jiang, B. Zhu, B. Xie (2020) Remote Sensing Images Data Augmentation Based on Style Transfer under the Condition of Few Samples. *Journal of Physics: Conference Series* **1653**(1) 012039.

[8] H. Wu, S. Zheng, J. Zhang, K. Huang (2019) Gp-gan: Towards realistic high-resolution image blending. Presented at *27th ACM International Conference.*

[9] A. Mikolajczyk, M. Grochowski (2018) Data augmentation for improving deep learning in image classification problem. Presented at *2018 International Interdisciplinary PhD Workshop* (IIPhDW).

[10] E.D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q.V. Le (2019) Autoaugment: Learning augmentation policies from data. Presented at *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR).

[11] M. Xu, S. Yoon, A. Fuentes, D.S. Park (2023) A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition* **137**(1) 109347.

[12] H. Inoue (2018) Data augmentation by Pairing Samples for images classification. Presented at *International Conference on Learning Representations* (ICLR 2018).

[13] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz (2018) Mixup: Beyond empirical risk minimization. Presented at *International Conference on Learning Representations* (ICLR 2018).

[14] L. Nanni, M. Paci, S. Brahnam, A. Lumini (2021) Comparison of different image data augmentation approaches. *Journal of Imaging* **7**(12) 254.

[15] D.V. Storozhik, A.G. Protasov (2022) Image processing technologies based on fusion data (Review). *Technical Diagnostics and Non-Destructive Testing* **4** 17-26.

[16] B. Wronski, I. Garcia-Dorado, M. Ernst, D. Kelly, M. Krainin, C. K. Liang, P. Milanfar (2019) Handheld multi-frame super-resolution. *ACM Transactions on Graphics (ToG)* **38**(4) 1-18.

[17] D.V. Storozhik, A.G. Protasov, O.V. Muraviov, V.F. Petrik, D.V. Petrenko (2022) Automation of thermal non-destructive testing process by applying the method of complexing thermographs. *Technical Diagnostics and Non-Destructive Testing* **2** 20-23.

[18] S.C. Wong, A. Gatt, V. Stamatescu, M.D. McDonnell (2016) Understanding data augmentation for classification: when to warp? Presented at *2016 International conference on digital image computing: techniques and applications* (IEEE DICTA-2016).

[19] A. Krizhevsky, I. Sutskever, G. Hinton (2012) Image Net classification with deep convolutional neural networks. Presented at *25th International Conference on Neural Information Processing Systems*.

[20] W.H. Kruskal, W.A. Wallis (1952) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47** 583-621.

[21] D.V. Storozhyk, O.V.Muraviov, A.G. Protasov, V.G. Bazhenov, G.A. Bohdan (2020) Multispectral image combining as a method of information content increasing at binary segmentation. *KPI Science News* **2** 82-87.